

LE FONCTIONNEMENT CONCURRENTIEL DU SECTEUR DE L'INTELLIGENCE ARTIFICIELLE GÉNÉRATIVE

Avis 24-A-05

Autorité
de la concurrence



DÉFINITION

- L'intelligence artificielle désigne tout outil utilisé par une machine « *afin de reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité* » (définition du Parlement européen)
- L'IA générative désigne des modèles d'IA capables de créer de nouveaux contenus qui peuvent par exemple être du texte, des images, du son ou des vidéos

UNE PRIORITÉ CROISSANTE DES POUVOIRS PUBLICS

Plusieurs initiatives :

- France : stratégie nationale pour l'IA, Commission de l'IA
- Europe : IA Act, DMA, Data Act
- Monde : priorité du G7, déclaration de Bletchley

Les bénéfices de l'IA générative ne se matérialiseront que si ménages et des entreprises ont accès à une diversité de modèles adaptés à leurs cas d'usage



CALENDRIER

- **8 février 2024** : Autosaisine d'office pour avis et lancement d'une consultation publique
 - Une quarantaine d'acteurs et une dizaine d'associations d'acteurs ont répondu au questionnaire
 - Entretiens avec des acteurs privés et internationaux du secteur et des acteurs institutionnels (services ministériels, autorité de régulation sectorielle, Commission européenne et autorités de concurrences étrangères)

- **29 mai 2024** : Séance de l'Autorité



OBJECTIFS ET PÉRIMÈTRE

PÉRIMÈTRE :

Avis concentré sur les stratégies mises en place par les grands acteurs du numérique visant à consolider leur pouvoir de marché à l'amont de la chaîne de valeur ou à tirer parti de ce pouvoir de marché pour se développer dans ce secteur en plein essor.

L'Autorité s'intéresse en particulier aux pratiques mises en œuvre par les acteurs déjà présents sur l'infrastructure d'informatique en nuage (cloud) et aux problématiques liées à l'accès à ces infrastructures, à la puissance de calcul, aux données et à une main d'œuvre qualifiée.

Elle examine également les prises de participation et les partenariats des grands acteurs du numérique, notamment dans des entreprises innovantes spécialisées dans l'IA générative.

OBJECTIFS :

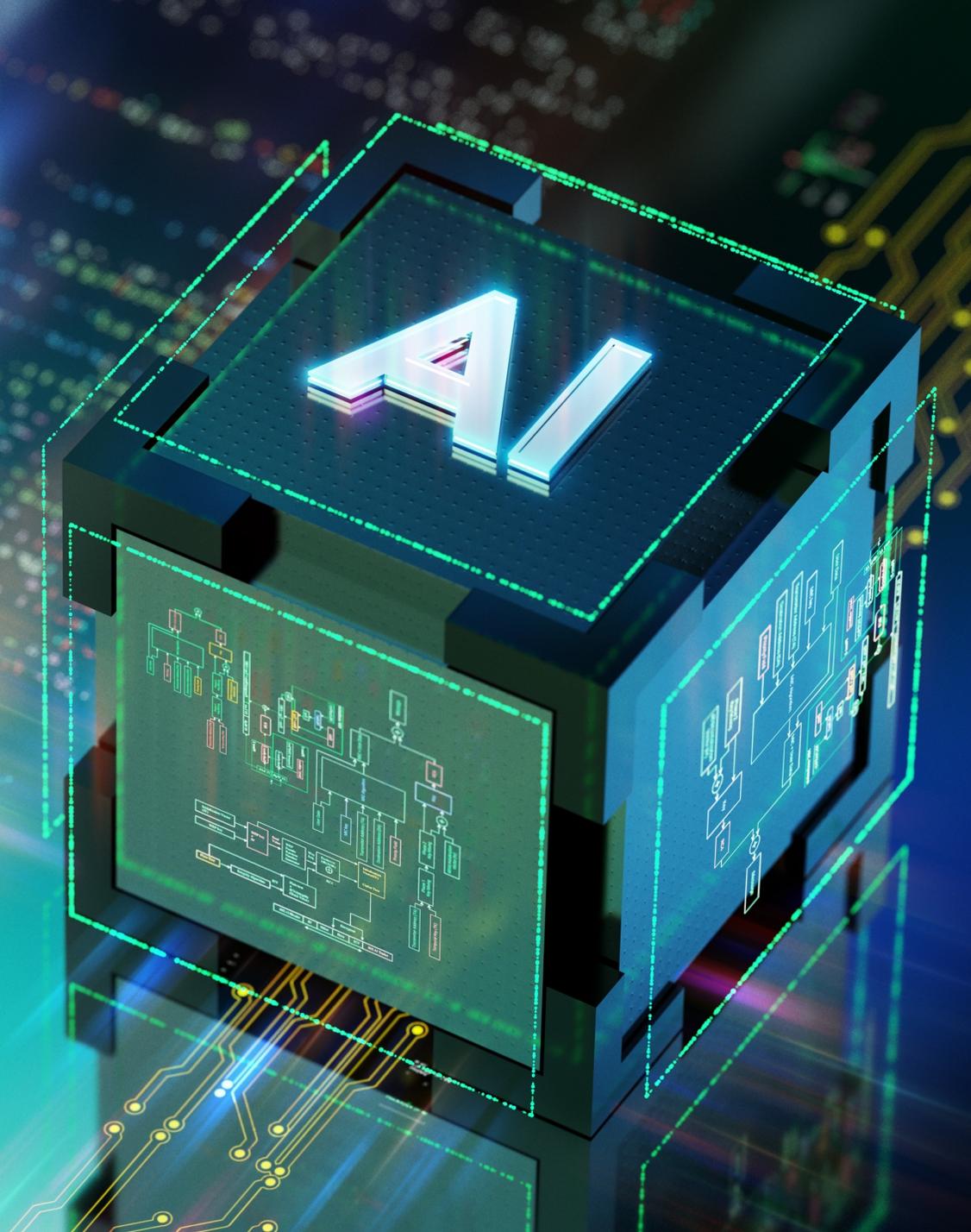
- Fournir aux acteurs une analyse du fonctionnement concurrentiel du marché de l'IA générative
- Formuler des recommandations visant à favoriser la dynamique concurrentielle du secteur



VALEUR AJOUTÉE DE L'AVIS

- Analyse de points structurants pour l'avenir:
 - Enjeux de la rémunération des contenus
 - Appréciation concurrentielle des participations minoritaires des grands acteurs au sein de jeunes entreprises innovantes
 - Levée d'une partie des barrières à l'entrée par la technologie open source
 - Pratiques susceptibles d'être mises en œuvre sur les marchés du travail
- Vigilance en ce qui concerne :
 - Les pratiques limitant abusivement l'accès à des intrants indispensables
 - Les partenariats conclus par des entreprises numériques déjà dominantes assortis ou pas de clauses d'exclusivité
 - Des pratiques de ventes liées susceptibles de consolider durablement le secteur de l'IA générative autour de ces entreprises





PLAN DE LA PRÉSENTATION

1. Le secteur de l'intelligence artificielle générative

- La chaîne de valeur
- Les acteurs de la chaîne de valeur

2. Le fonctionnement concurrentiel du secteur

- Les barrières à l'entrée
- Les avantages de certains acteurs liés à leur présence sur d'autres marchés
- Les risques concurrentiels

3. Les recommandations de l'Autorité



PARTIE 1

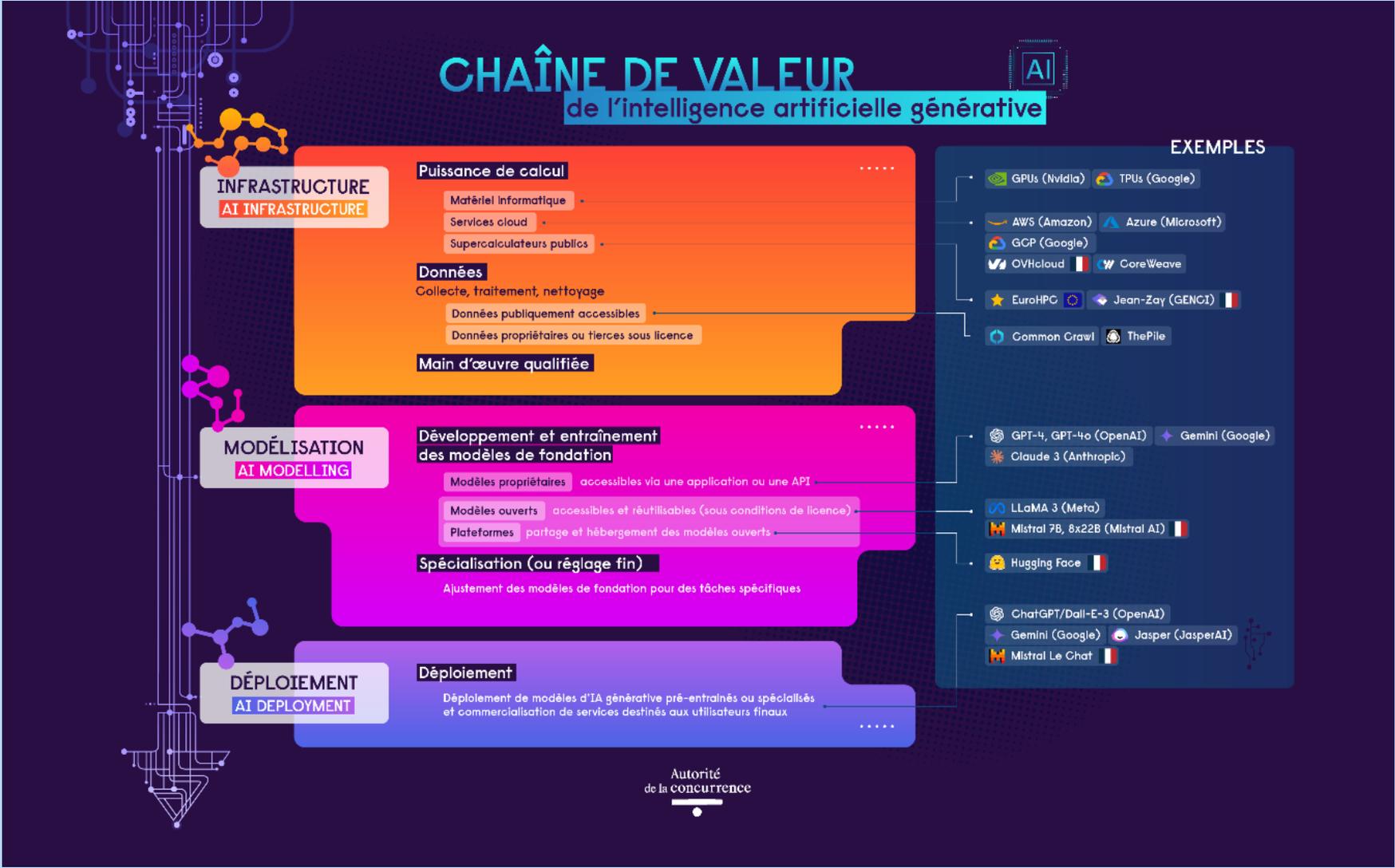
Le secteur de l'IA généralive

LA CHAÎNE DE VALEUR

Deux phases essentielles de la modélisation de l'IA générative sont à différencier :

- **L'entraînement** désigne le processus d'apprentissage initial d'un modèle à l'occasion duquel ses paramètres, appelés « poids », seront déterminés. Cet entraînement nécessite à la fois une grande puissance de calcul et un volume important de données, généralement publiques.
 - Il peut être complété par une phase de **spécialisation**, aussi appelée réglage fin, pendant laquelle le modèle va être adapté à une tâche particulière. La spécialisation se fait généralement à partir d'un ensemble plus petit de données propriétaires, et peut faire appel à l'expertise humaine
- **L'inférence** correspond à l'utilisation du modèle, une fois entraîné, pour créer du contenu. Le modèle peut être mis à disposition des utilisateurs via une application (Le Chat de Mistral AI) ou une interface de programmation pour les développeurs. La puissance de calcul nécessaire dépend du nombre d'utilisateurs.
 - L'inférence peut également être complétée par l'apport de nouvelles données (qui n'ont pas été utilisées pour l'entraînement), pour ancrer le modèle (« *grounding* »), par exemple dans l'actualité la plus récente.

LA CHAÎNE DE VALEUR DE L'IA GÉNÉRATIVE



LES ACTEURS DE LA CHAÎNE DE VALEUR

- Les **géants du numérique** : Alphabet et Microsoft sont intégrés verticalement et de manière conglomérale sur toute la chaîne de valeur, tandis qu'Amazon, Apple, Meta et Nvidia sont seulement présents sur certaines couches spécifiques
- Les **développeurs de modèles** comme Anthropic, HuggingFace Mistral AI ou OpenAI. Ces développeurs ont souvent noué des partenariats avec un ou plusieurs géants du numérique.

À l'amont, plusieurs types d'acteurs sont présents:

- Les **fournisseurs de composants informatiques** développent des processeurs graphiques (GPU) et des accélérateurs d'IA. Nvidia est le principal acteur du secteur.
- Les **fournisseurs de services cloud** comprennent à la fois des « *hyperscalers* » comme Amazon Web Services (AWS), Microsoft Azure et Google Cloud Platform (GCP), des fournisseurs de *cloud* comme OVHCloud ainsi que des fournisseurs spécialisés dans l'IA comme CoreWeave.

À l'aval, de nombreux acteurs développent de nouveaux services fondés sur l'IA générative à destination du grand public (comme ChatGPT), des entreprises et des acteurs publics et/ou intègrent l'IA générative dans leurs services existants (comme Zoom).



PARTIE 2

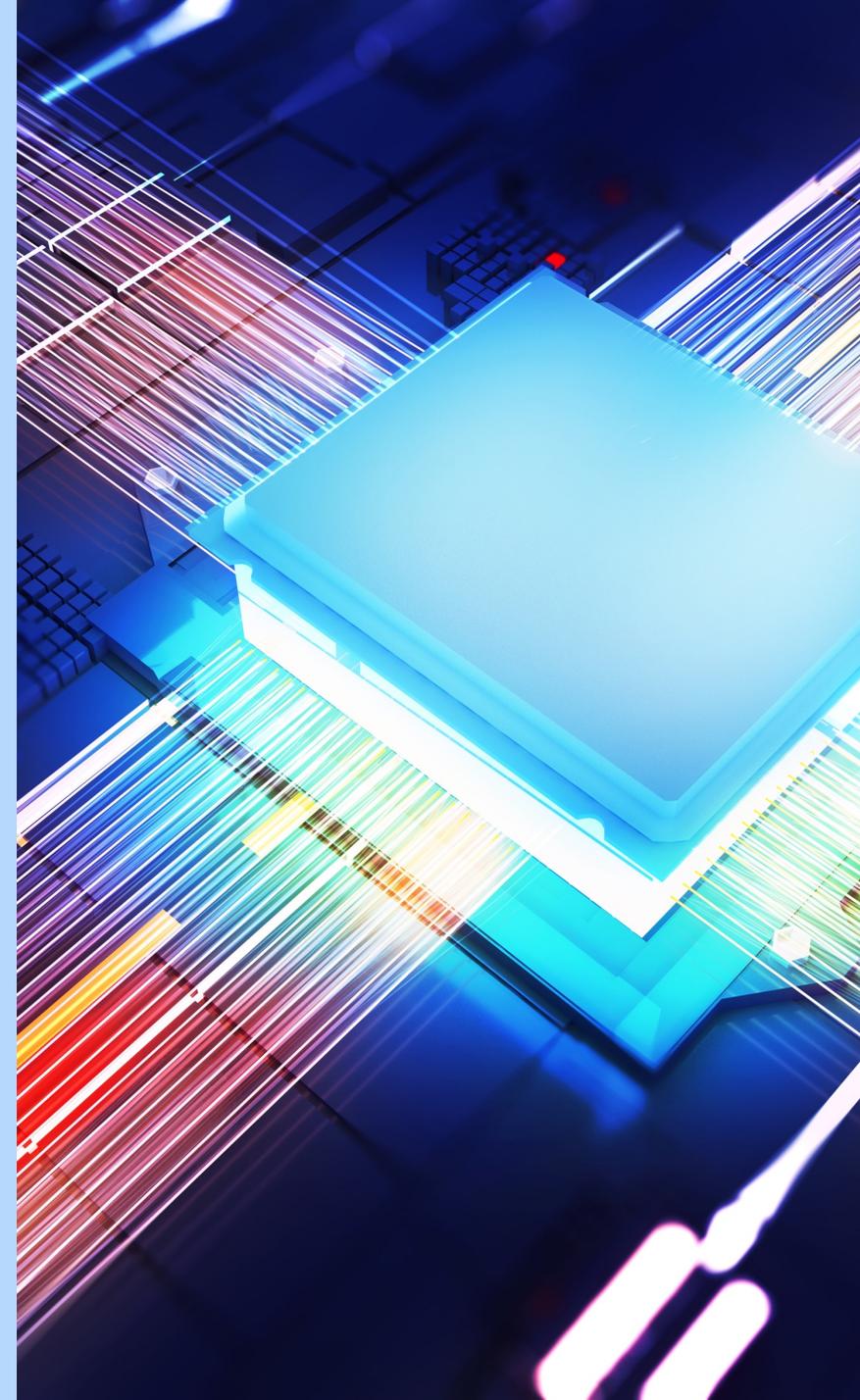
Le fonctionnement concurrentiel du secteur

DES BARRIÈRES À L'ENTRÉE ELEVÉES

- L'accès à une puissance de calcul suffisante pour effectuer un grand nombre d'opérations en parallèle et dotée d'une forte précision est essentiel
- Le *cloud* est la solution privilégiée pour l'entraînement ou la spécialisation des modèles et permet également de faciliter le déploiement à l'aval
- L'entraînement des modèles nécessite un vaste ensemble de données
- Des compétences rares et très recherchées
- Besoins de financements multipliés par six entre 2022 et 2023

FACTEURS SUSCEPTIBLES DE LIMITER CES BARRIÈRES À L'ENTRÉE

- Le recours aux supercalculateurs publics (Jean Zay en France)
- Certaines innovations technologiques permettent de réduire le besoin en données et en puissance de calcul (données synthétiques)
- L'existence de modèles ouverts



AVANTAGES DE CERTAINS ACTEURS LIÉS À LEURS ACTIVITES SUR D'AUTRES MARCHÉS

L'accès privilégié aux intrants

Les grandes entreprises du numérique ont un accès privilégié aux intrants nécessaires pour l'entraînement et le développement des modèles de fondation.

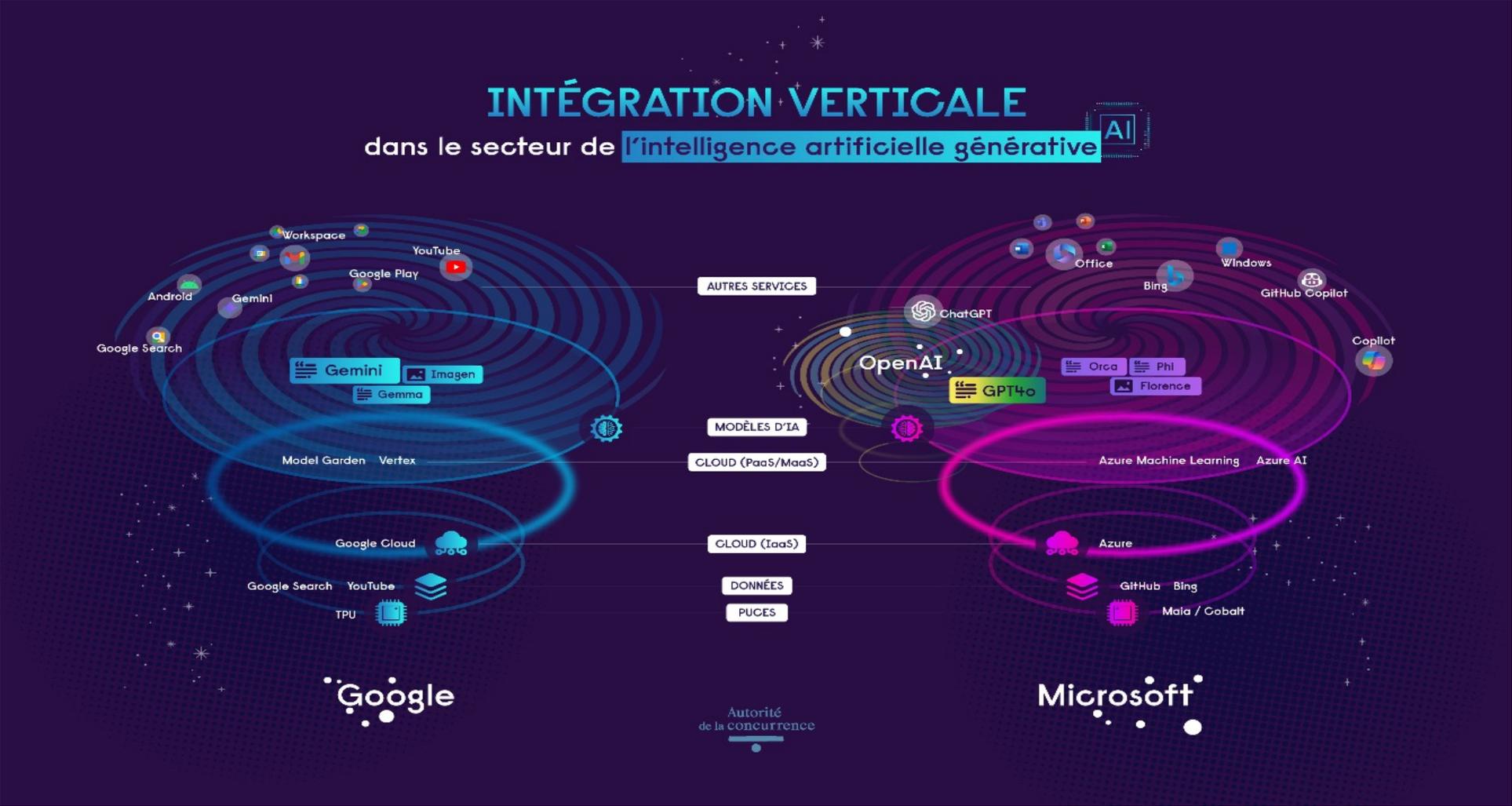
- Un accès facilité à la puissance de calcul
- Un accès privilégié à un large volume de données (comme Alphabet avec YouTube). Elles ont aussi la faculté d'utiliser leur puissance financière pour conclure des accords avec des fournisseurs de contenus.
- Un grand nombre de talents sont attirés par les salaires attractifs et les conditions de travail offerts au sein des grandes entreprises du numérique.

Les avantages découlant de leur intégration verticale et conglomérale

- Économies d'échelle, de gammes et effets de réseaux
- Ces entreprises commencent à intégrer les outils d'IA générative dans leurs écosystèmes de produits et de services (Microsoft Copilot)
- Les places de marché de modèles (MaaS) de ces grandes entreprises permettent d'accéder à des modèles d'IA générative, propriétaire ou de tiers, conçus pour fonctionner dans leur écosystème *cloud*



LA POSITION DE CERTAINS ACTEURS SUR LES MARCHÉS CONNEXES



LES RISQUES CONCURRENTIELS (1)

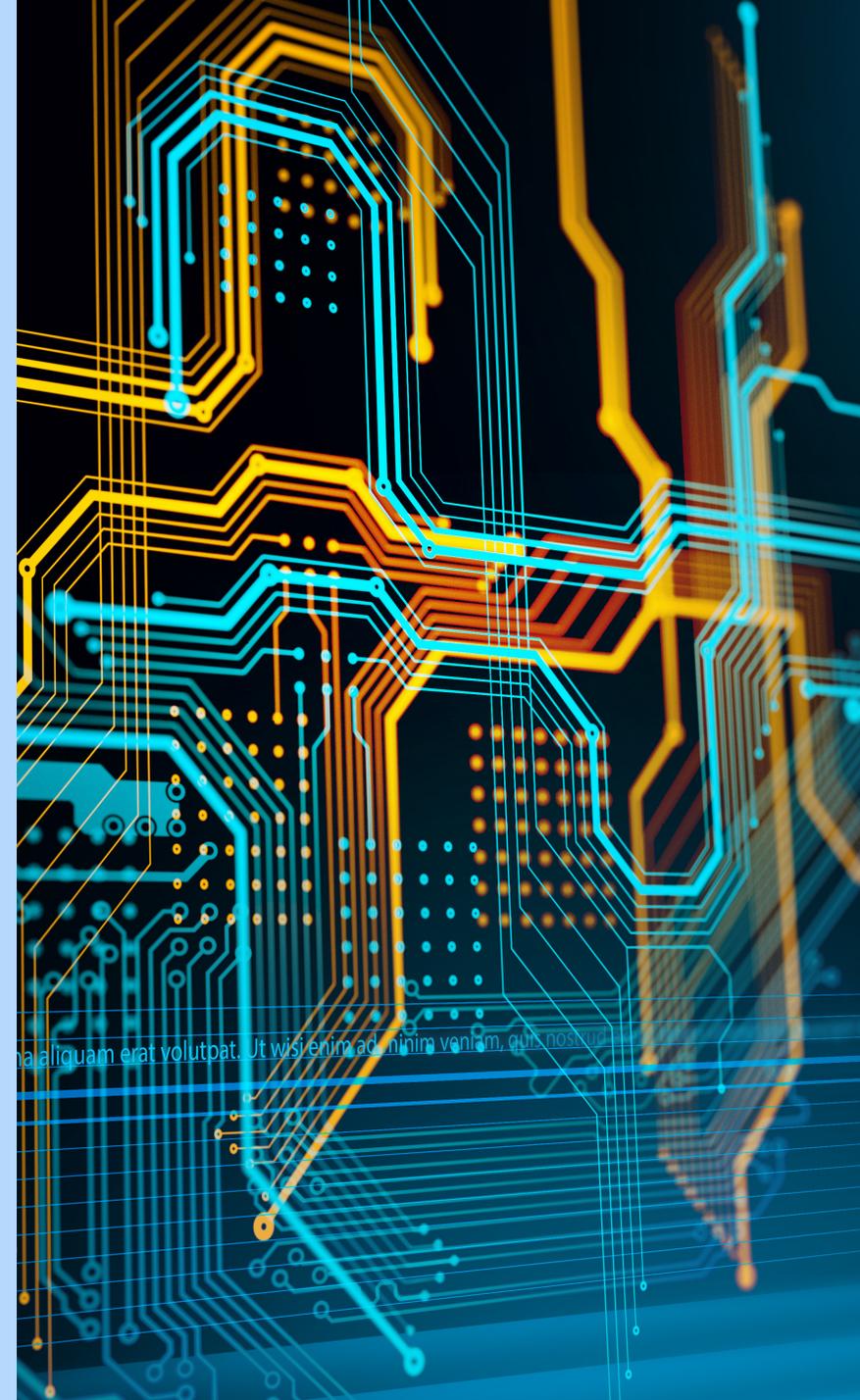
Difficile à ce stade de définir des **marchés pertinents**. Dans certains cas, l'analyse devra se faire en termes d'**écosystèmes** plutôt que marché par marché.

L'Autorité identifie plusieurs risques d'abus :

- **Au niveau des composants informatiques** : fixations de prix, restrictions d'approvisionnement, conditions contractuelles déloyales ou comportements discriminatoires. Également des préoccupations liées au logiciel CUDA
- **Un risque de verrouillage par les grands fournisseurs de services *cloud*** : offres de crédits *cloud* élevées à destination spécifiquement des jeunes entreprises innovantes du secteur et pratiques de verrouillage technique

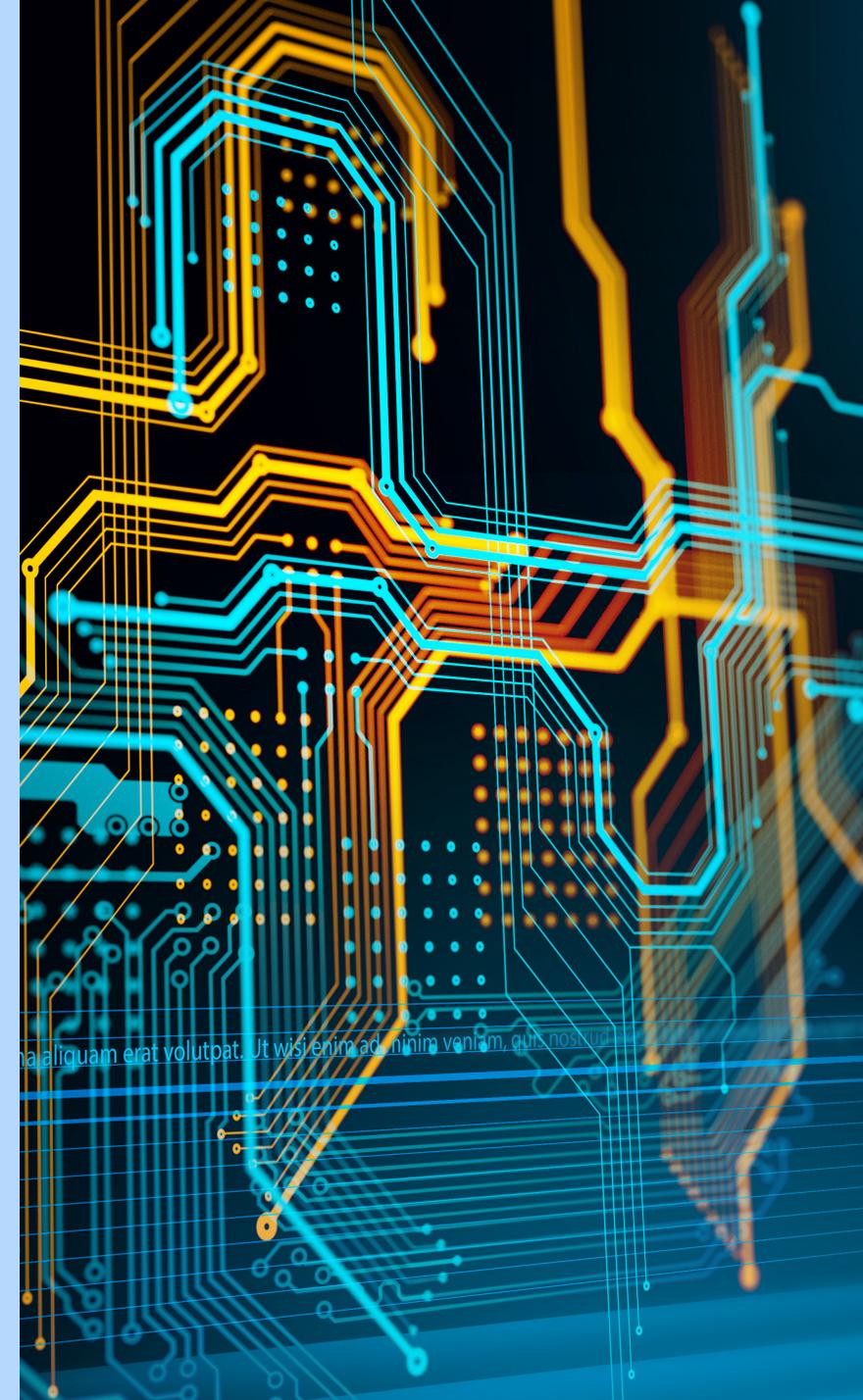
L'Autorité identifie **plusieurs préoccupations de concurrence concernant l'accès aux données**, notamment :

- Les accords par lesquels de grandes entreprises du numérique **verseraient des rémunérations importantes** difficilement répliquables par leurs concurrents
- L'usage abusif de règles relatives au **RGPD** limitant l'accès aux données
- L'exploitation des contenus des éditeurs par les fournisseurs de modèles de fondation, **sans l'autorisation des détenteurs de droits**.



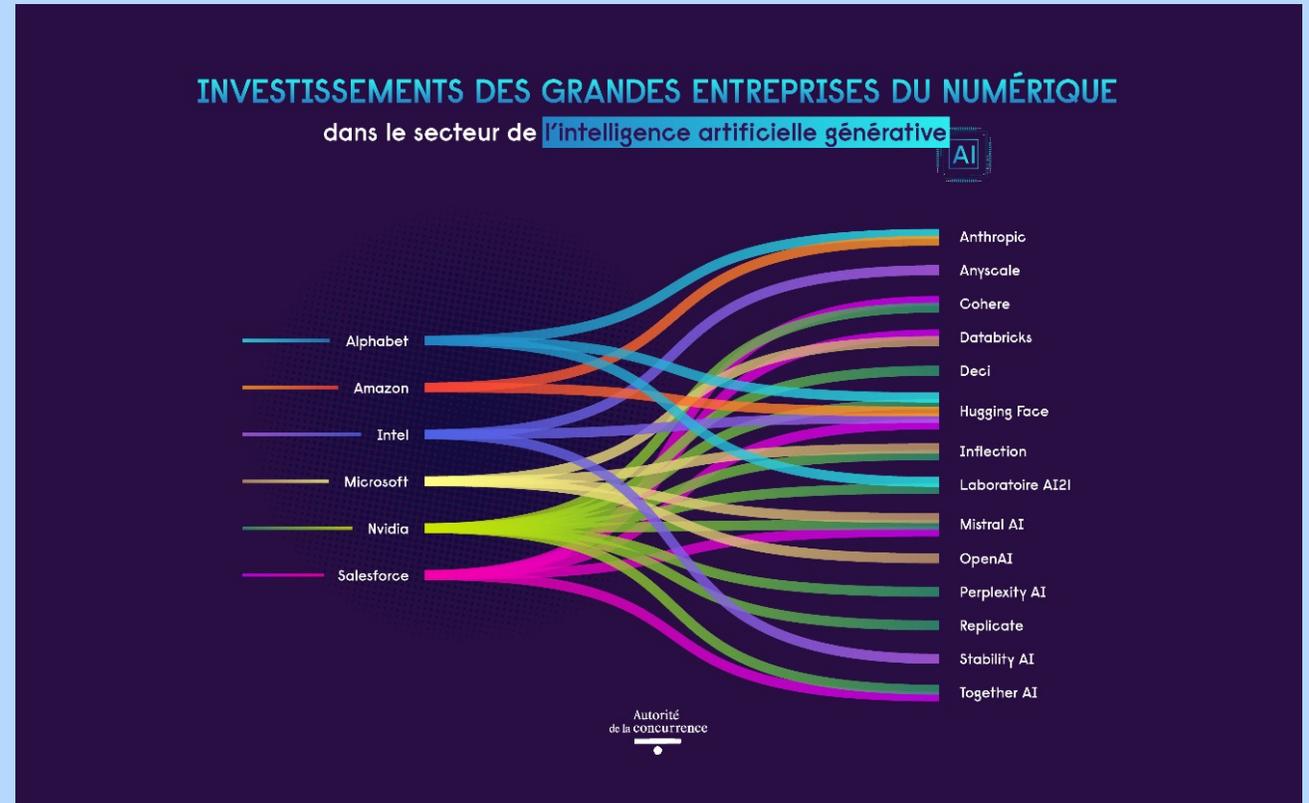
LES RISQUES CONCURRENTIELS (2)

- **Les risques liés à l'accès à une main d'œuvre qualifiée**
 - Les **accords de non-débauchage**
 - **Le recrutement, par les géants du numérique d'employés stratégiques** (exemple du recrutement par Microsoft d'une grande partie des 70 employés de la *start-up* Inflection) soulève aussi des préoccupations
- **Les risques liés aux modèles en accès libre:** les conditions d'accès et de réutilisation des modèles ou de certains de leurs composants peuvent conduire au verrouillage des utilisateurs (précédent « Google Android »)
- **Les risques liés à la présence d'entreprises sur plusieurs marchés distincts**
 - À l'amont, des risques de refus ou des limites d'accès à des puces ou des données nécessaires pour entraîner des modèles de fondation concurrents ainsi que des accords d'exclusivité
 - **A l'aval, l'intégration d'outils d'IA générative sur certains supports, comme les *smartphones*,** suscite notamment des inquiétudes (exemple de l'accord Google/Samsung)
- Risques de pratiques concertées



LES RISQUES CONCURRENTIELS (3)

- Les prises de participations minoritaires et partenariats des géants du numérique peuvent soulever des préoccupations de concurrence
- **Nécessité d'une vigilance compte tenu des risques concurrentiels** : affaiblissement de l'intensité concurrentielle, effets verticaux, transparence du marché, verrouillage de certains acteurs
- Ces opérations peuvent être appréhendées par les autorités de concurrence sur plusieurs fondements
- L'Autorité constate toutefois **un manque de transparence** sur ces accords



PARTIE 3

Les recommandations émises par l'Autorité

RENDRE PLUS EFFICACE LE CADRE RÉGLEMENTAIRE EXISTANT

Compte tenu de la rapidité avec laquelle le secteur évolue, ces réglementations devront être pleinement mises en œuvre et leur impact évalué afin d'éviter des effets négatifs sur l'innovation et la concurrence.

À droit constant, des améliorations pourraient cependant être envisagées.

Proposition n°1 : la Commission devrait porter une attention particulière au développement des services MaaS pour évaluer la possibilité de désigner les entreprises fournissant de tels services en tant que contrôleurs d'accès au titre du DMA.

Proposition n°2 : dans la mise en œuvre des dispositions de la loi SREN sur les avoirs d'informatique en nuage, la DGCCRF devrait accorder une attention particulière à l'utilisation de ces avoirs dans le domaine de l'IA.

Proposition n°3 : le futur Bureau de l'IA, mis en place à l'article 64 du règlement sur l'IA, et l'autorité nationale compétente en France, qui sera désignée en application de l'article 70 du règlement sur l'IA, devront s'assurer d'une part que la mise en œuvre du règlement ne freine pas l'émergence ou l'expansion d'opérateurs de taille plus modeste, et d'autre part que les plus grands acteurs du secteur ne détournent pas le texte à leur avantage.



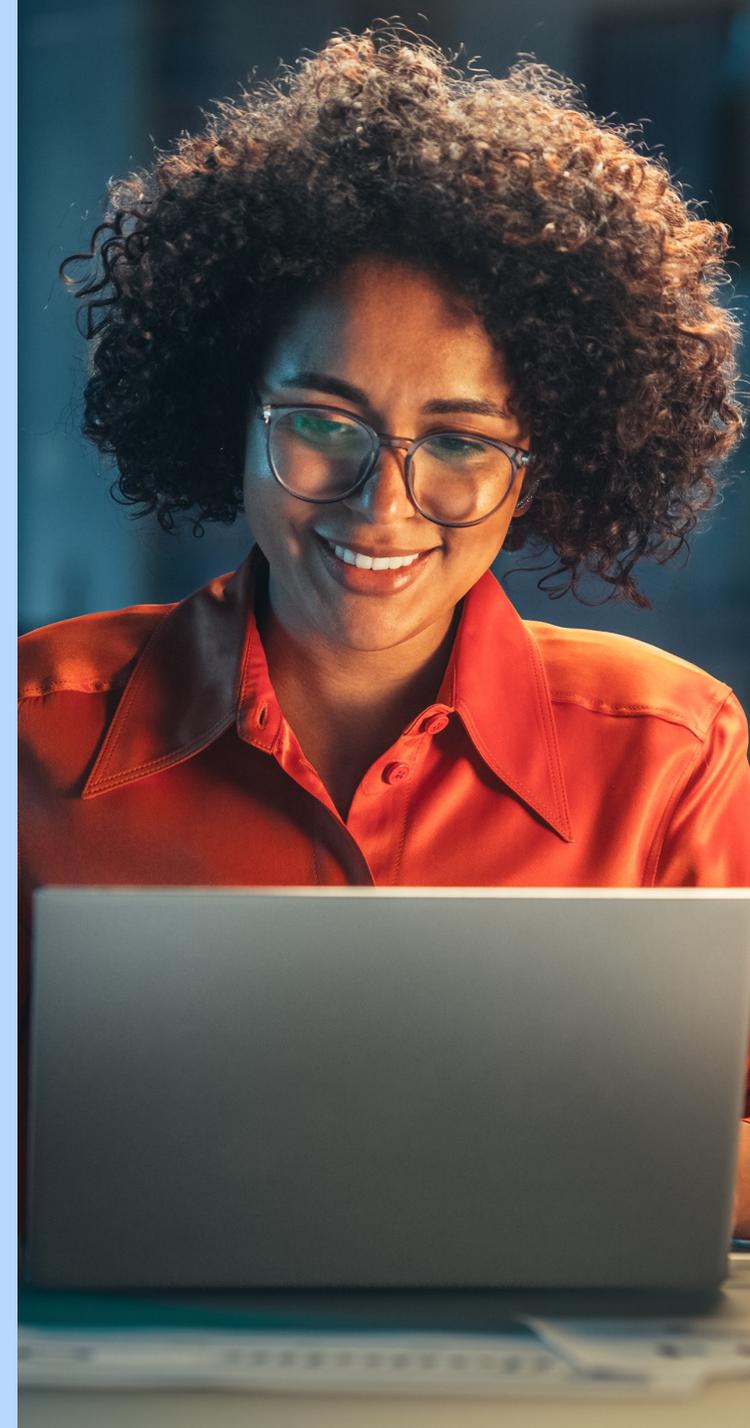
MOBILISER LES OUTILS DU DROIT DE LA CONCURRENCE

La mobilisation des outils de concurrence, y compris du point de vue des pratiques restrictives de concurrence relevant principalement de la DGCCRF, aura un rôle essentiel pour prévenir l'émergence ou la consolidation de positions dominantes ou des ententes qui affecteraient la dynamique concurrentielle du secteur

Les services de l'Autorité peuvent s'appuyer sur plusieurs initiatives :

- Étude conjointe avec le Bundeskartellamt sur les algorithmes (2019)
- Avis sur le fonctionnement concurrentiel du secteur du *cloud* (2023)
- Opérations de visites et saisies dans le secteur des cartes graphiques (2023)
- Analyse du manquement au devoir d'information des ayant-droits dans l'affaire Google « droits voisins » (2024)

Proposition n°4 : les autorités chargées de la régulation concurrentielle des marchés devront demeurer vigilantes dans le secteur de l'IA générative et mobiliser, si nécessaire, l'ensemble des outils à leur disposition pour agir de manière rapide et efficace.



DÉVELOPPER L'ACCÈS À LA PUISSANCE DE CALCUL

L'accès à la puissance de calcul est indispensable pour permettre le développement de la recherche et l'émergence de nouvelles entreprises dans le secteur de l'IA générative.

Proposition n°5 : poursuivre les investissements dans le développement des supercalculateurs au niveau européen, pour permettre au plus grand nombre d'acteurs d'accéder à la puissance de calcul.

Proposition n°6 : le gouvernement et/ou les sociétés assurant la gestion des supercalculateurs pourraient engager une réflexion afin de proposer un cadre ouvert et non-discriminatoire permettant à des acteurs privés d'utiliser les ressources des supercalculateurs publics contre rémunération, tout en conservant la priorité aux recherches notamment académiques.

Proposition n°7 : en lien notamment avec l'AI Act, fixer des critères d'ouverture des modèles d'IA génératives entraînés sur des supercalculateurs publics.



TENIR COMPTE DE LA VALEUR ÉCONOMIQUE DES DONNÉES

Les usages de la donnée diffèrent selon le stade auquel elles sont utilisées au sein de la chaîne de valeur (entraînement, réglage fin ou inférence). Les accords entre ayants-droits et développeurs devraient refléter l'importance relative de ces données pour les développeurs. Par ailleurs, les coûts de transaction sont un point d'attention.

Proposition n°8 : les autorités publiques, notamment dans le cadre de la mission confiée par la Ministre de la culture au Conseil supérieur de la propriété littéraire et artistique, pourraient inciter les ayants droits à tenir compte de la valeur économique des données selon les cas d'usage (en introduisant par exemple des prix différenciés), et à proposer des offres groupées réduisant les coûts de transactions, ceci afin de garantir les capacités d'innovation des développeurs de modèles

Proposition n°9 : faciliter la mise à disposition des données de la sphère publique et privée pour l'entraînement ou le réglage fin de modèles d'IA générative, et encourager les initiatives publiques ou privées visant à diffuser les données francophones, qu'il s'agisse de textes, d'images ou de vidéos.



ASSURER UNE MEILLEURE TRANSPARENCE SUR LES PRISES DE PARTICIPATIONS DES GÉANTS DU NUMÉRIQUE

Dans l'attente de la décision de la Cour européenne de justice sur l'article 22 du règlement sur les concentrations, le cadre juridique actuel permet d'appréhender la plupart des préoccupations de concurrence concernant les accords entre entreprises que cela soit *ex ante* par le droit du contrôle des concentrations ou *ex post* par le droit des pratiques anticoncurrentielles (ententes ou abus de position dominante).

Néanmoins, à droit constant une meilleure transparence des participations minoritaires non contrôlantes dans le secteur serait justifiée.

Proposition n°10: A l'occasion de l'obligation d'information des concentrations prévue à l'article 14 du DMA, la Commission pourrait également demander, dans le modèle relatif à l'article 14 du règlement sur les marchés numériques, des informations sur les participations minoritaires détenues dans le même secteur d'activité que la cible.

